Volume 12 • Number 8 • 2009
VALUE IN HEALTH

ISPOR Health Policy Council Proposed Good Research Practices for Comparative Effectiveness Research: Benefit or Harm?

Til Stürmer, MD, MPH, Tim Carey, MD, MPH, Charles Poole, MD, ScD

Department of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC, USA

There are increasing calls for better understanding of "what works" in health care [1]. One of the means for assessing what works is through "comparative effectiveness research" (CER) [2]. Ideally, the needed data would come from randomized controlled trials (RCTs) or from natural experiments. RCTs would need to be large, practical clinical trials that compare interventions head-to-head in real clinical settings [3,4], using novel approaches to assess clinically relevant outcomes.

Nonexperimental studies of intended drug effects have been criticized because confounding by indication (selective channeling of patients to treatment modalities based on outcome predictors such as severity of disease) can almost never be ruled out [5,6]. Recent developments in pharmacoepidemiologic methods limit the potential for bias, and thus increase the value of nonexperimental comparisons of intended drug effects. These methods include instrumental variable methods [7,8], the new user design [9], the use of a comparator drug with a similar indication to that of the index drug [10], propensity scores [11], and simple improvements such as eliminating immortal person-time [12] and reducing selection bias by not censoring follow-up when a person stops taking a drug [13,14]. Much remains to be done, however, including the study of heterogeneity of treatment effects at the intersection between personalized medicine and pharmacoepidemiology. In addition, there remains an unresolved tension between emulating RCTs (increasing internal validity based on increasing restrictions [15]) and enhancing generalizability (external validity).

CER is an interdisciplinary endeavor in which the disciplines are linked by the need for information and the development of methods. The involvement of ISPOR in this enterprise is welcome. Like drugs, the Good Research Practices proposed by the working group of the ISPOR Health Policy Council and published in this issue of the journal [16–18] need to be evaluated by their potential benefits and harms. The potential benefits are obvious. Someone unfamiliar with performing non-experimental comparisons of drugs and their outcomes will find valuable discussion in these documents of issues to be considered. Common to all such documents, however, there is the potential for harm when the recommendations are used as a cookbook without understanding their interplay. References to standard textbooks of pharmacoepidemiology [e.g., 19] could help alleviate this problem.

Address correspondence to: Til Stürmer, Department of Epidemiology UNC Gillings School of Global Public Health Epidemiology, McGavran Greenberg Hall, CB # 7435, Chapel Hill, NC 27599-7435, USA. E-mail: til.sturmer@post.harvard.edu

10.1111/j.1524-4733.2009.00653.x

Dr. Tim Carey has no conflict to declare; he receives salary support from the Agency for Healthcare Research and Quality.
Dr. Charles Poole receives salary support from the UNC-GSK Center of Excellence in Pharmacoepidemiology and Public Health and has received speaking fees from GSK.

It is inevitable for any detailed overview to contain questionable or outmoded recommendations. Recommendations of the proposed documents that some experienced pharmacoepidemiologists might find arguable include:

1. The requirement to report the results from all *ex ante* analyses. Some such analyses will have been abandoned because the researchers discovered that they are biased.
2. The assessment of the importance of biases based on how they affect the acceptance or rejection of the null hypothesis. Biases are best measured by their effects on the magnitude, direction and precision of effect-measure estimates.
3. The recommendation for propensity score models to include variables that are only weakly related to treatment selection (but unrelated to the outcome per the following recommendation). It is unclear why any variable that is unrelated to the outcome should be included in a propensity score [20].

Some core issues in the design and analysis of non-experimental comparisons are not addressed in enough detail. One is the importance of the role of various "stakeholders" in CER. Another is the distinction between confounding (e.g., by indication) and selection bias (due, for instance, to non-adherence, drop-out of "sick stoppers," etc. [13]). The potential to separate these forms of bias is one of the main advantages of the new user design [9].

Given the expected continuation of the rapid development of pharmacoepidemiologic methods over the past 5 years, the proposed Good Research Practices may become outdated very rapidly [21]. We found no indication of how ISPOR intends to keep these guidelines up to date. In an era of guideline proliferation, one might ask what the proposed ISPOR document will add to the existing ones in this field, especially the Good Pharmacoepidemiologic Practice document published and continuously updated by the International Society for Pharmacoepidemiology (ISPE) [22]. Finally, harmonization of the ISPOR documents with others, including those proposed by ISPE, the US Institute of Medicine, the US Agency for Healthcare Research and Quality, the UK National Institute for Health and Clinical Excellence should be considered. Such harmonization would prevent confusion and nit picking by groups opposed to CER.

In our view, the benefit-to-harm balance of ISPOR's proposed documents on Good Research Practices favors the benefit side. It will help to spread the news that non-experimental treatment comparisons are possible given careful design, analysis, and interpretation. We congratulate ISPOR for providing guidelines for CER that emphasize the potential benefits without giving CER a black box warning for its potential harms.

Source of financial support: Dr. Til Stürmer receives investigator-initiated research funding and support as principal investigator and coinvestigator from the National Institute on Aging at the National Institutes of Health. He also receives research funding as coprincipal investigator of the

  1098-3015/09/1042 1042–1043

UNC-DEcIDE center from the Agency for Healthcare Research and Quality. Dr. Stürmer does not accept personal compensation of any kind from any pharmaceutical company, though he receives salary support from the UNC-GSK Center of Excellence in Pharmacoepidemiology and Public Health and from unrestricted research grants from pharmaceutical companies to UNC.

References

1 Eden J, Wheatley B, McNeil B, Sox H, eds. Knowing What Works in Health Care: A Roadmap for the Nation. Washington, DC: National Academies Press, 2008.
2 Institute of Medicine. Committee on Comparative Effectiveness Research Prioritization, Board on Health Care Services. Initial national priorities for comparative effectivness research. Washington, DC: National Academies Press, 2009.
3 Tunis S. Comparative Effectiveness: Basic Terms and Concepts. San Francisco, CA: Center for Medical Technology Policy, 2007. Available from: http://www.allhealth.org/briefingmaterials/Tunis4-27-07-699.pdf [Accessed September 8, 2009].
4 Kolata G. New arena for testing of drugs: real world. *NY Times*; November 25, 2008.
5 Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2:267–71.
6 Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? Am J Med 1984;77:703–8.
7 Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722–9.
8 Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 2006;17:268–75.
9 Ray W. Evaluating medication effects outside of clinical trials: new-user designs. Am J Epidemiol 2003;158:915–20.
10 Glynn RJ, Schneeweiss S, Stürmer T. Indications for propensity scores and review of their use in pharmacoepidemiology. Basic Clin Pharmacol Toxicol 2006;98:253–9.
11 Stürmer T, Schneeweiss S, Brookhart MA, et al. Analytic Strategies to adjust confounding using Exposure Propensity Scores and Disease Risk Scores: Nonsteroidal Antiinflammatory Drugs (NSAID) and Short-Term Mortality in the elderly. Am J Epidemiol 2005;161:891–8.
12 Suissa S. Immortal time bias in pharmaco-epidemiology. Am J Epidemiol 2008;167:492–9.
13 Andersen M, Brookhart MA, Glynn RJ, et al. Practical issues in measuring cessation and re-initiation of drug use in databases. Pharmacoepidemiol Drug Saf 2008;17(Suppl. 1):S27.
14 Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology 2008;19:766–79.
15 Schneeweiss S, Patrick A, Stürmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. Med Care 2007;45(Suppl. 2):S131–42. Supplement 2: emerging methods in comparative effectiveness and safety.
16 Berger ML, Mamdani M, Atkins D, Johnson ML. Good research practices for comparative effectiveness research: defining, reporting and interpreting non-randomized studies of treatment effects using secondary data sources: the ISPOR retrospective database analysis task force report—Part I. Value Health 2009; in press.
17 Cox E, Martin BC, Van Staa T, et al. Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of non-randomized studies of treatment effects using secondary data sources: the ISPOR retrospective database analysis task force report—Part II. Value Health 2009; in press.
18 Johnson ML, Crown W, Martin BC, et al. Good research practices for comparative effectiveness research: analytic methods to improve causal inference from non-randomized studies of treatment effects using secondary data sources. The ISPOR retrospective database analysis task force report—Part III. Value Health 2009; in press.
19 Strom BL, ed. Pharmacoepidemiology (4th ed.). Chichester: John Wiley & Sons Ltd., 2005.
20 Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. Am J Epidemiol 2006;163:1149–56.
21 Rothman KJ, Poole C. Some guidelines on guidelines: they should come with expiration dates. Epidemiology 2007;18:794–6.
22 ISPE. Guidelines for good pharmacoepidemiology practices (GPP). Pharmacoepidemiol Drug Saf 2008;17:200–8.